Can a Chatbot Increase the Motivation to Provide Personal Health Information?

Kerstin DENECKE [a,1], Richard MAY [a], Annkathrin PÖPEL [b],
Sandra LUTZ HOCHREUTENER [c]
[a] *Bern University of Applied Sciences, Bern, Switzerland*
[b] *Sanatorium Kilchberg, Zurich, Switzerland*
[c] *Zurich University of Arts, Zurich, Switzerland*

Abstract. In healthcare settings, questionnaires are used to collect information from a patient. A standard method for this are paper-based questionnaires, but they are often complex to understand or long and frustrating to fill. To increase motivation, we developed a chatbot-based system Ana that asks questions that are normally asked using paper forms or in face-to-face encounters. Ana has been developed for the specific use case of collecting the music biography in the context of music therapy. In this paper, we compare user motivation, relevance of answers and time needed to answer the questions depending on the data entry method (i.e. app Ana versus paper-based questionnaire). A randomised trial was performed with 26 students of music therapy. The results show that the chatbot is more motivating and answers are given faster than on paper. No differences in answer relevance could be determined between the two means. We conclude that a chatbot could become an additional data entry method for collecting personal health information.

Keywords. Conversational user interface, medical history, self-reported clinical data, mHealth

Introduction

In medical settings, paper-based questionnaires are used to collect relevant data directly from the patient. Computer-assisted history taking systems are tools that aim to aid physicians in gathering data from patients to decide for a diagnosis, a treatment plan or both [1]. These systems have potential of collecting more correct data (less falsified by patients) [2], in particular when it comes to sensitive questions such as alcohol consumption [3]. Quaak et al. showed that computerised history taking is suitable for certain patients in addition to, and not as a substitute for, the oral interview with a doctor [4]. Often implemented as digitized questionnaires [5], these systems cannot motivate the patients to answer an often comprehensive catalogue of medical history questions [6].

In previous work, we developed Ana, a mobile system for collecting information on the music biography of a patient [7]. The music biography is of interest for therapists delivering music therapy. Music therapists incorporate a range of music making methods within and through a therapeutic relationship [8]. As part of this relationship, the music biography of a person is relevant to decide for an appropriate interaction with the patient, to find out which experiences (good and bad) a patient made with music to select means for the therapy. Ana enables a patient to think about the questions on his or her music biography, to enter the answers while communicating with the chatbot and finally to send

---

[1] Corresponding Author: Kertin Denecke, Bern University of Applied Sciences, Quellgasse 21, Biel, Switzerland; E-mail: kerstin.denecke@bfh.ch

the replies to the therapist after completing. Information collected comprises aspects on the current behaviour with respect to music making, listening and moving to music as well as the past music biography including effects music had on a patient. Ana is implemented with a conversational user interface (CUI). The CUI is rule-based, exploiting the Artificial Intelligence Markup Language (AIML) for encapsulating questions and responses of the system. Questions are asked one after another and the user can reply either using free text or by selecting one out of several options. In case some question is unclear a user can ask for explanations. More details are presented in Denecke et al. [7]. The usability of Ana had been studied already with students with technical background who not necessarily had a comprehensive music biography [7]. Following this evaluation, detected usability issues were resolved and the set of questions was adapted.

The objective of this work is to assess whether the chatbot Ana can help motivating users in delivering information on themselves, or on their music biography respectively. To facilitate interaction with mHealth applications and increase adherence to the digital intervention, chatbots or CUI have been applied in last years [9]. A CUI-based system is a computer program that interacts with users using natural language (written or spoken). The aim of such a system is to simulate a human conversation. While some CUI-based systems allow free text responses by users, others restrict the user input to selecting specific predefined items (e.g. choosing options as replies). Some systems use embodied avatars, while others reduce the conversation to an exchange of text messages as known from messengers such as WhatsApp. Health chatbots may increase user engagement and usability [10] and may also solve issues with healthcare service availability. In particular for the treatment of mental disorders, chatbots gain in interest to address shortages and difficulties in accessing mental health services [11]. Other application areas in health care include: medication management [12] or triage systems [13]. Often it is unclear whether chatbot systems are at least as efficient as the existing means to deliver a health care service, or in our case, to collect the music biography. In this paper, we assess the efficacy of Ana in terms of motivation and time for providing personal health information in a randomised trial and by analysing the given answers.

1. Methods

We designed a randomised trial where participants were asked to answer questions on their music biography using 1) a paper-based questionnaire or 2) the app Ana. The following research questions were assessed:

- **Answer relevance:** Is there a difference in response relevance and degree of detail of responses depending on the data entry method (paper vs. app)?
- **Time**: Is there a difference in time needed to fill the questionnaire (paper vs. app)?
- **Motivation**: Are people rather willing to answer additional questions on their music biography when using the app or the paper-based questionnaire?

The study population was recruited from students of the Zurich University of Arts. In this way, we could ensure that the participants have a music biography and they really have information to provide on their background with music.

The time needed to answer the questions was measured and recorded by two persons. Relevance of answers was manually assessed by a music therapist on a scale of 0 (irrelevant) to 5 (extremely relevant). Comprehensibility of answers was quantified by the number of words per answer. Feedback on the motivation was collected using a feedback questionnaire. To test the differences in the number of words used to answer the question (i.e. reflecting the comprehensibility of responses) regarding statistical significance, we performed a Welch Two Sample t-Test with the following hypotheses:

- H0: Average number of words in answers given in the app-based music biography equals or is smaller than the average number of words in answers of the paper-based questionnaire.
- H1: Average number of words in answers given in the app-based music biography is larger than the average number of words in replies in the paper-based questionnaire

The instrument (app, paper) to be used by a participant was distributed randomly. Since the evaluation took place during the regular lessons of the students, the time for the test was limited to 30 minutes in total. The participants had 15-20 minutes for interacting with the app and the remaining time to provide feedback on a separate feedback questionnaire. The number of questions integrated in Ana had to be reduced to 33 for the evaluation due to a time slot of 30 minutes available for the evaluation. We basically removed questions dealing with early experiences with music since we believe that it is more time consuming to think about early childhood experiences. The questions asked by the app and to be filled on paper were the same. All participants joined a brief introductory session where the app was explained. A pre-test was conducted with 13 students of medical informatics to ensure that all questions on the music biography can be answered within approximately 15 minutes and sufficient time is left for filling the feedback questionnaire. Two feedback questionnaires had been created: one to be filled by the participants who had to use the paper-based questionnaire and one for those using Ana. The questions concerned motivation of the participant, perspicuity of the questions in the app or on the paper questionnaire, subjective impressions whether the questions could have been answered faster using the other medium. The feedback questionnaire for the paper-based music biography comprised 7 statements to be answered on a 5-item Likert scale (see fig. 2). The feedback questionnaire for the app comprised 10 statements to be answered by selecting one value on a 5-item Likert scale (see fig. 1), and one additional free text query on possible improvements. The Cantonal Ethics Committee confirmed in a clarification of responsibility that for the planned trial no ethics approval is required.

Table 1. Results of the randomised trial (n=13 per group)

	Average age (range)	Average time needed	Male / Female	Average number of words	Relevance (scale 0-5)
Group 1 (App)	39.62 (27-53)	19 minutes	38.5% / 61.5%	4.96	4.71
Group 2 (Paper)	44.38 (23-58)	21 minutes	15.4% / 84.6%	5.6	4.75

2. Results

During the randomised trial, 13 persons filled the music biography using the app (group 1) and another 13 people replied using the paper-based questionnaire (group 2), see Table 1. The participants in group 1 were slightly younger in average and also the distribution of female and male differed among the two groups. In total, more women participated in the trial. The time needed to fill the questionnaire was smaller for group 1. Relevance of answers to free text queries was high with 4.71 (group 1) and 4.75 (group 2) and similar for both groups. Comprehensibility of replies is manifested by the number of words used to answer. In total, participants of group 1 used 104 words to answer all the free text queries. In contrast, 117 words in total are used in average by participants in group 2. The null hypothesis cannot be rejected (p=0.8897): the difference in the number of words is not significant.
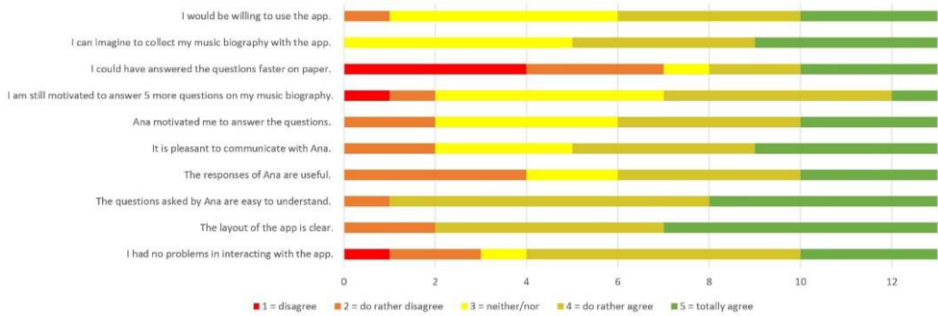

Figure 1. Feedback after answering questions on music biography using the app Ana (n=13)
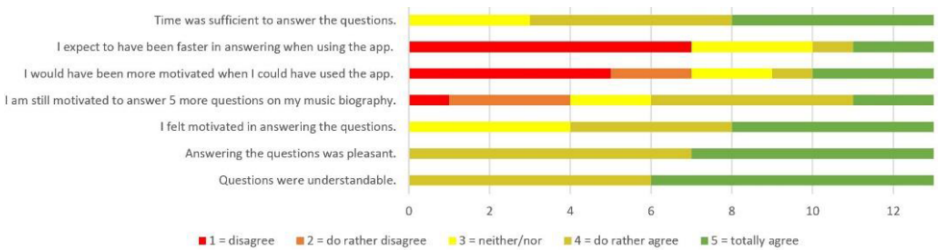

Figure 2. Feedback after answering questions on music biography using a paper-based questionnaire (n=13).

From the feedback questionnaires with results shown in Fig. 1 and 2 we can see the following: The majority of participants who used the app would be willing to use the app in future and can imagine to collect the music biography using the app. Questions were rated as understandable and the layout was clear. 61.5% of the app users confirmed that it is pleasant to communicate with Ana. Regarding motivation, 7 persons (53.8%) claimed that the app motivated them to answer the questions; 4 persons (30.8%) are neutral in their judgement regarding motivation. Similarly, 7 persons (53.8%) claimed they would be still motivated to answer additional questions using Ana after finishing; 4 persons (30.8%) are neutral. Compared to the results from the participants who filled the paper-based music biography, we can see that 4 persons (30.8%) were no longer

motivated to answer additional questions on the music biography. In the app-group, only two persons were no longer motivated. While all persons in the group 2 agreed that the questions are understandable, in group 1 one person disagreed with that statement. Only 4 persons (30.8%) of group 2 claim that they would have been more motivated when using the app instead of paper. They made this statement without having a chance to have a look at the app. 7 persons (53.8%) believe that they would not have been faster in filling the music biography when using a paper-based questionnaire.

3. Discussion

From the results of the randomised trial, we can conclude that there is neither a relevant difference in the degree of detail in the answers nor in the relevance of the answers among the two types of instrument. Chatbot or paper questionnaire could be used interchangeably. Time needed to answer the questions is slightly higher when using the paper-based instrument. Since the individual music experience might differ among the participants, this could also impact on the length of the replies (having less experiences would result in shorter replies). However, since all participants had a background in music, we expect that the answers of both groups are comparable in our randomised trial and that it is really faster to answer the questions using the app. Regarding motivation to answer the questions, it seems that Ana rather motivates to answer the questions than the paper-based questionnaire. Brandtzaeget al. found out that productivity, entertainment and social motivations are the main motivational factors for people to use chatbots [14]. We did not assessed the reasons why people claimed they would be willing to answer additional questions. We can only assume that entertainment or productivity were also reasons that impacted the motivation in our study.

Besides increased motivation in the patients, a benefit is that the answers are digital available and users can ask Ana for explanations in case they do not understand a question. The age span of the participants was quite broad, but did not included persons older than 60. Thus, we cannot conclude that older users will also feel comfortable in using the app. Further, study participants were not patients: perceptions might be different when persons suffer from a mental disease or other disease. Typing during the evaluation was impacted by using an unknown mobile phone. Wrong language settings of the phone (set to English instead German) led to errors in spelling correction. During the evaluation, we recognised some technical issues with the chatbot: The system posted the same question several times which frustrated the users. Since the bot currently does not store the context, it misinterpreted song titles which led to unexpected behaviour of the bot. For example, the German children song "Guten Abend, Gute Nacht" (Good evening, good night) was recognised as greeting by the chatbot who restarted the entire conversation from the beginning and questions had to be answered again.

For some responses no adequate answer was given by Ana. We have to note here that Ana is simply collecting the responses and is not interpreting the answers of a user. However, our users were sometimes irritated from human-like behaviour. The participants criticized that Ana was encouraging and complimenting them. While recent research considers humanizing chatbots [15], it seems that the users in our context see the human-likeness rather critically. This aspect has to be considered carefully in future development to ensure that Ana (or any other health chatbot) is implemented in a way that it is accepted by users.

4. Conclusion

Our comparative experiment (paper versus chatbot) provided first hints that a chatbot can motivate more in providing personal health information than a paper questionnaire. We see the benefits of a chatbot for asking questions that normally have to be answered using paper-based questionnaires or in face-to-face encounters: chatbot technologies can be designed specifically to be linguistically and culturally adaptive. Thus, Ana could be implemented in a way to adapt to the health literacy of a user or to explain questions when users do not understand at the first instant. Currently experts in music therapy assess the questions included in Ana for their practical relevance and a study with real patients receiving music therapy is planned. Such evaluation is necessary to confirm our results in a real world setting. Additionally, Ana's knowledge base is adapted to be applicable to the field of general medicine. In this context, the app will be used to collect information in preparation to the first encounter between physician and patient.

References

[1]   Vaira L, Bochicchio MA. Smart Anamnesis for Gyn-Obs: Issues and Opportunities. In: Longo A et al., editors. Cloud Infrastructures, Services, and IoT Systems for Smart Cities, Springer, 2017, 95–104.

[2]   Lauritsen K, Degl'Innocenti A, Hendel L, Præst J, Lytje MF, Clemmensen-Rotne K and Wiklund I, Symptom recording in a randomised clinical trial: paper diaries vs. electronic or telephone data capture. Controlled clinical trials 2004; 25,6: 585–597.

[3]   Carr AC, Ghosh A, Ancill, R. Can a computer take a psychiatric history? Psychological Medicine. 1983; 13(1): 151–158.

[4]   Quaak MJ, Westerman RF and van Bemmel JH. Comparisons between written and computerised patient histories. Br Med J (Clin Res Ed). 1987; 295,6591: 184–190.

[5]   Zakim D. Development and significance of automated history-taking software for clinical medicine, clinical research and basic medical science, Journal of internal medicine. 2016; 280,3: 287–299.

[6]   Slack WV, Kowaloff HB, Davis RB, Delbanco T, Locke SE, Safran C and Bleich HL. Evaluation of computer-based medical histories taken by patients at home, J Am Med Inform Assoc. 2012; 19,4: 545–548.

[7]   Denecke K, Lutz Hochreutener S, Pöpel A, May R. Self-Anamnesis with a Conversational User Interface: Concept and Usability Study. Methods Inf Med. 2018 Nov;57(5-06): 243-252.

[8]   Bunt L. Music therapy: An art beyond words. Routledge, 2003.

[9]   Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, Surian D, Gallego B, Magrabi F, Lau AJS, Coiera E. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc. 2018; 25,9: 1248–1258. doi:10.1093/jamia/ocy072.

[10] Fadhil A and Gabrielli S. Addressing challenges in promoting healthy lifestyles: the al-chatbot approach. In: Proceedings of the 11th EAI international conference on pervasive computing Technologies for Healthcare, ACM, 2017; p. 261–265.

[11] Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Can J Psychiatry. 2019;64,7:456-464.

[12] Tschanz M, Dorner TL, Holm J and Denecke K, Using eMMA to manage medication, Computer. 2018; 51,8:18–25.

[13] Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a chatbot (Ada) in the diagnosis of mental disorders: comparative case study with lay and expert users. JMIR formative research. 2019; 3(4), e13863.

[14] Brandtzaeg PB, Følstad A. Why people use chatbots. In: International Conference on Internet Science, Springer, 2017, 377–392.

[15] Go E and Sundar SS. Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. Computers in Human Behavior. 2019; 97: 304–316.